Standard Operating Procedures for the Command Data Model and Command Data Dictionary

Chapter 1 - Introduction

Implementing Information Technology (IT) that fully supports the mission of the organization, specifically providing quality services and information to customers, remains a strategic vision for the US Army Corps of Engineers (USACE). This vision encompasses leveraging USACE strategic business with Information Technology initiatives to truly recognize the benefits and return on investment from major programs and projects organization-wide.

An objective of the Directorate of Information Management (DIM) is to ensure that all functional areas share data to increase the accuracy, reliability, and usability of information. In addition, much emphasis has been focused on innovative techniques and methodologies of data management, data standardization, and data reuse as a guiding principle for system development efforts.

The Command Data Model (CDM) and the Command Data Dictionary (CDD) serve as the framework to enforce data sharing, data standards, and interoperability. This principle not only applies to the development and acquisition of IT investment projects, but also to the management of the organization through discrete business rules and consistent terms and definitions documented in the model. The CDM is a TO-BE conceptual data model that represents the USACE corporate view. The CDD contains the definitions of the CDM data entities and data elements, as well as supporting characteristics about each data element, such as character type, length, domain values, and domain descriptions. This type of data is referred to as metadata, meaning data about data. The models integrated in the CDM form the basic framework for information management of all business and systems initiatives that envision sharing data.

1-1. Purpose and Objectives

The purpose of this Engineering Publication (EP) is to provide users, functional proponents, and others with a step-by-step approach for utilizing, maintaining, and implementing the CDM and CDD. The primary objectives of this EP are:

- Facilitate a common understanding of how using the CDM/CDD adds value to the organization.

- Ensure maximum consistency and shareability of data, both within USACE and appropriate external activities.

- Promote the use of standardized data elements within and across business and functional processes.

- Facilitate functional user involvement in information systems development.

- Enforce use and maintenance of the CDM.

1-2. Scope

The scope of this EP focuses on the USACE CDM and CDD and how the two components support corporate information management and systems development projects.

1-3. Document Composition

This EP is divided into the following chapters:

CHAPTER 1. INTRODUCTION -- provides background, purpose, and objectives of this EP.

CHAPTER 2. FUNDAMENTAL CONCEPTS -- presents an overview of USACE standard methodology and techniques to facilitate the use and maintenance of the CDM.

CHAPTER 3. STANDARD OPERATING PROCEDURES (SOP) -- provides standard operating procedures on how to use, implement, and maintain the CDM USACE-wide.

CHAPTER 4. SUPPORTING TOOLS -- provides an overview of the Corps' Data Encyclopedia and summarizes the tools that are used to facilitate the phases documented in section 3.2. Each section will provide the purpose and major objectives of each tool.

1-4. Applicability

This EP applies to the following components:

- USACE/Office of the Chief of Engineers (OCE) elements, Major Subordinate Commands (MSC), Districts, Laboratories, and Field Operating Activities (FOA)

- USACE information systems and subsystems

- USACE Data Management Activities

1-5. Roles and Responsibilities

Data Architecture Control Committee (DACC). The DACC will support the Information Resources Management Steering Committee (IRMSC) in planning an orderly transition to a shared data environment. The DACC will establish logical and physical data structures with associated data definitions that promote corporate system integration, data sharing, data standardization, data integrity and interoperability. The DACC will also manage and enforce the development and maintenance of corporate AISs as embodied in the CDM, including cross-functional processes, data integration, and life cycle management of data.

Command Data Administrator (CDA). As specified in the AR 25-1 and the DoD Directive (DODD) 8320.1, the CDA will:

(1) Plan, oversee, and enforce policies and procedures governing Data Administration activities. This includes coordinating and managing USACE-wide standards, guidelines, and procedures pertaining to data administration activities.

(2) Manage and advocate the use and maintenance of the Command Data Model (CDM). This function includes promoting IDEF0 and IDEF1X modeling, supporting functional areas in model development efforts based on the Corps' corporate model, and ensuring that the CDM data structures are up-to-date, based on interaction with functional area proponents, and captured in the USACE Data Encyclopedia. The CDA is also responsible for data integration activities to include impact analysis, model correctness checks, and three-schema correlations.

(3) Enforce data standards and naming conventions. Based on USACE, Department of Army (DA) and DoD Data Standardization procedures, the CDA interfaces with functional area proponents and system developers to ensure that the "standard" USACE data is used as a baseline for systems development efforts. This function includes ensuring the use of standard prime words, class words, and proper definitions as they relate to accomplishing the USACE business mission.

(4) Enforce the use of USACE data administration procedures, tools, and methodologies. The CDA assists functional proponents with using the CDM and Data Encyclopedia to identify critical success factors for a successful systems development project.

(5) Coordinate issues and actions with the USACE Command Database Administrator (CDBA). As conceptual models are developed, the CDA interacts with the CDBA in developing physical models. This function includes promoting data sharing and integration, ensuring that data integrity is maintained, and validating the accuracy of system designs and development in accordance with functional area business models and the Life Cycle Management of Information Systems (LCMIS) guidelines.

(6) Coordinate proposed changes to data objects received from the functional proponent with DACC and IRMSC for corporate resolution. The CDA serves as the liaison between functional proponents and system developers. As new data objects and/or overlapping data objects are discovered, the CDA facilitates the process of presenting proposals and recommendations to the DACC for consensus and resolution. As solutions are reached, the CDA ensures updates are documented in the CDM as required. The CDA also ensures that formal documentation and approval procedures for all changes to data objects, metadata, and data definitions are enforced.

Command Database Administrator (CDBA). The CDBA will establish the definition, organization, protection, control, and efficiency of the shared physical database structure, ensuring data is defined and organized to support multiple users. The CDBA will also establish guidelines for division, district and applications DBAs. The CDBA coordinates with the CDA and with other DBAs. The CDBA must ensure that principles of Command and Control-Protect (C2-Protect) Strategy and Risk management are employed to identify and manage risks in order to protect information in the AIS.

Data Stewards. The data steward is assigned by the HQUSACE Directors and Chiefs of separate offices to define functional data requirements for AISs. The person or group will manage the development, approval, creation, and use of data associated with specific data standards managed within a specified functional business area.

Functional Proponents. Functional Proponents (FP) are responsible for compliance with this SOP within their area of responsibility. The FP will designate a data administrator within their area of responsibility who will serve as the key point of contact in implementation of data administration policies and practices. The FP data administrator coordinates data administration activities with the data steward.

Systems Developers/Operations Managers. Systems Developers/Operations Managers will ensure compliance with the Data Administration policy for developing, implementing, and managing AISs. The designated command, agent, or office is assigned mission area responsibility for all AISs under technical development, acquisition, modification, and/or maintenance to coordinate with the assigned data administrator within their business area..

1-6. References

1. Department of Defense (DoD) Directive 8320.1, “DoD Data Administration,” September 26, 1991

2. DoD Manual 8320.1-M, “Data Administration Procedures,” March 1994

3. DoD Manual 8320.1-M-1, “Data Standardization Procedures,” March 1996

4. Federal Information Processing Standards Publication (FIBS PUBS) 183, Integration Definition for Function Modeling (IDEF0), December 21, 1993

5. Federal Information Processing Standards Publication (FIBS PUBS) 184, “Specifications for Integration Definition for Information Modeling (IDEF1X),” December 21, 1993

6. “Defense Data Repository System (DDRS) End User Manual,” September 1, 1994

7. “Department of the Army Technical Architecture, “ Version 4.0, January 30, 1996

8. Army Regulation 25-1, “The Army Information Resources Management Program,” March 25, 1997

9. Army Regulation 25-3, “Army Life Cycle Management of Information Systems,” October 15, 1989

10. Engineer Regulation 25-1-2, “Life Cycle Management of Automated Information Systems”, November 30, 1991

11. “Manager’s Guide to Life Cycle Management of Automated Information Systems,” March 1994

12. Engineer Regulation 25-1-100, “Data Administration Program,” (draft)

13. USACE Strategic Data Management Plan, June 1992

14. Data Encyclopedia End User Manual, October 18, 1991

1-7. Glossary

Appendix I provides the glossary of terms and definitions.

Chapter 2 - Fundamental Concepts

This chapter presents an overview of USACE standard methodology and techniques to facilitate the use, implementation, and maintenance of the CDM. Successful development and implementation of a shared data environment embraces fundamental concepts and techniques. These techniques offer the flexibility to document business requirements, graphically illustrate business rules, and migrate to IT solutions. This chapter will address frequently asked questions such as:

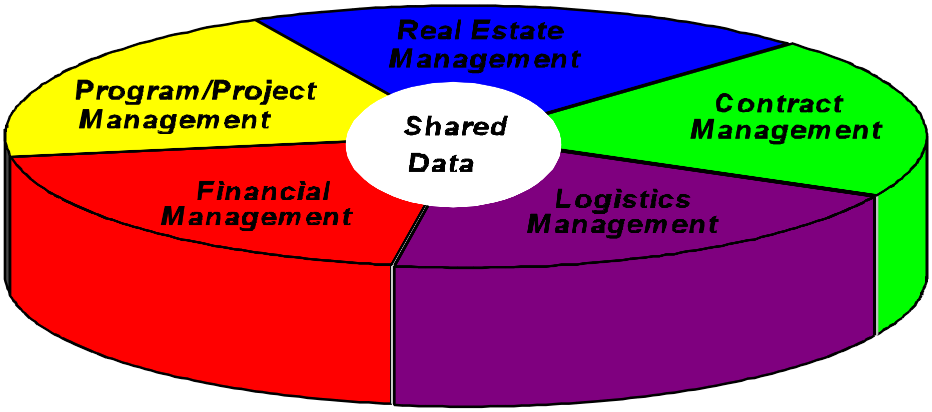
- What is the Command Data Model and Command Data Dictionary?
- What are the benefits of using a data model and data dictionary such as the CDM/CDD?
- How does the CDM relate to the DDM?
- What data modeling methodology and techniques can help achieve business objectives?
- What are standard naming conventions?
- What is three schema architecture?
- How do conceptual, physical, and external schemas relate to one another?

2-1. Overview of the Command Data Model (CDM)

History of the Command Data Model

The CDM is a TO-BE data model which contains the data entities, data elements, business rules, domains, data definitions, data proponency, and other metadata for all USACE business areas. The CDM was originally developed from three of the original USACE baseline application development concepts (ADPs): Work-Breakdown-Structure, Work-Item-Resource-Structure, and Accounting Structure. The “baselines” provide USACE managers with a framework for segregating work into smaller segments (projects, sub-projects, and tasks), assigning resources to work areas, and accounting for appropriations and expenditures related to work. The baselines incorporate the most highly shared data in the CDM, thus the CDM is the foundation for conceptual model integration initiatives throughout the USACE as shown in Figure 2-1.

Figure 2-1. CDM Shared Data



The CDM is composed of shared data across functional areas which has been reviewed and approved by the DACC prior to integration. The CDM is comprised of integrated data from Real Estate Management, Contract Management, Logistics Management, Financial Management, Program/Project Management and other business areas. The business rules within these areas are subsumed in the CDM, and changes to one business area are reflected throughout the other areas. An extract of the integrated data models for each business area are periodically taken from the CDM.

The CDM is stored in the Data Encyclopedia, the Corps' data repository, to facilitate data sharing, development, maintenance and technical support of USACE information systems as reference in Section 4-1, "Overview of the Data Encyclopedia".

Program managers, functional proponents, and developers recognize the importance of data exchange and have established guiding principles to facilitate data integration, data standardization, and data sharing across USACE. These guiding principles are:

- Data must be maintained separate from the supporting application;
- Data is a corporate resource (asset);
- Data should be easily accessible and shared by all;
- The minimum essential data required to meet functional business needs must be maintained;
- User involvement in development is critical, and;

- Continuous improvement information quality, timeliness, and access is essential.

Data and information are the two most critical strategic resources used by USACE for planning and implementing information systems. The CDM gives the functional user a standard reference for data definitions, data formats, and business rules. Data accuracy is improved because of improved access, better definition, and the maintenance of additional attributes about the data. The result is reduced time and cost correcting data errors and processing rejects.

2-2. Overview of the Command Data Dictionary (CDD)

The CDD is a glossary of all data contained in the CDM which can be retrieved online via the USACE Data Encyclopedia repository. The CDD is the authoritative source for data standards and definitions. The dictionary contains registered data with related metadata and provides documentation of the life cycle events for standard data. The dictionary is a reporting mechanism comprised of reports of various business areas which are integrated in the CDM.

2-3. Benefits

The benefits of using and maintaining the CDM/CDD include sharing information and improving the efficiency and effectiveness of business decisions. From a business process perspective, managers and functional proponents will recognize the following benefits:

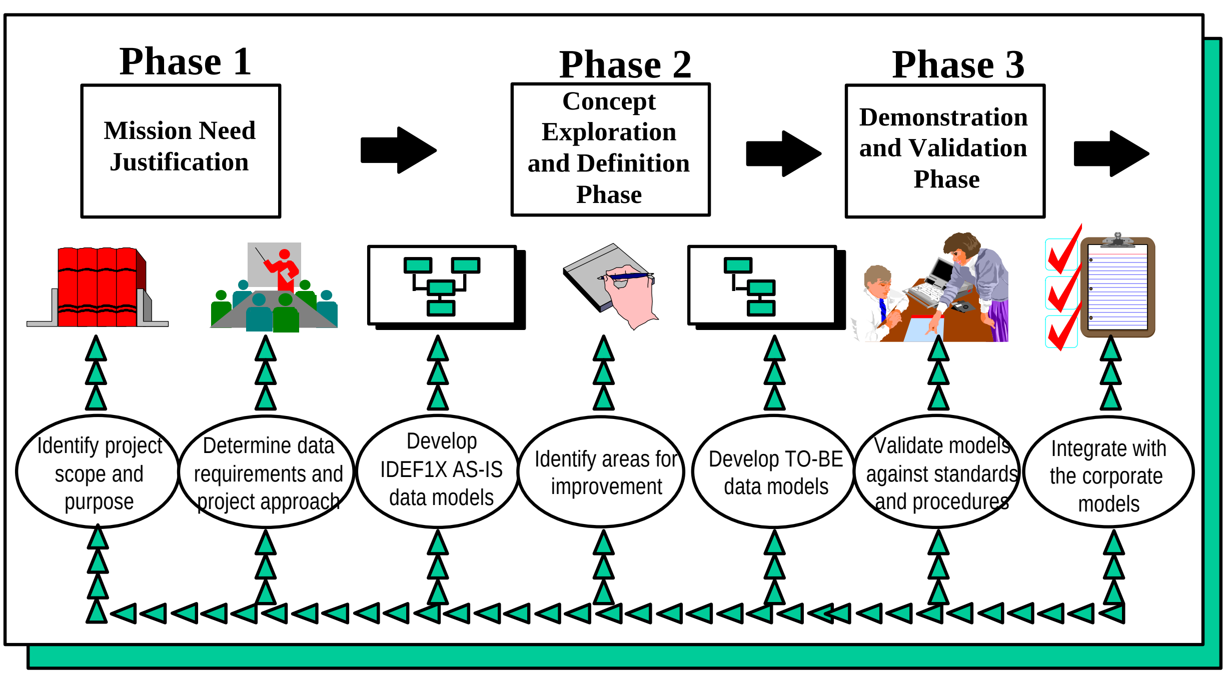
- Reduced data redundancy;
- Reuse of data and software asset;
- Improved data accuracy;
- Improved timeliness of data;
- Facilitated business area analysis and improvement projects, and;
- Documentation for the USACE information architecture.

From a systems development perspective, the CDM and CDD yield the following benefits:

- Better conformance with other USACE, Army, and DoD Data Architecture;
- Identification of reusable software components;
- Fewer data errors which reduce maintenance costs to retrofit;
- Identification of requirements for software acquisition and comparison;
- Physical models that can be automatically generated;
- Documentation for systems extracted from the CDM and the CDD;
- Impact analysis performed across the conceptual framework, physical database views, and external view of the data, and;
- Functional requirements in support of the Life Cycle Management of Information Systems (LCMIS) milestone requirements.

There are five phases of the LCMIS process. They include: (1) Mission Need Justification, (2) Concept Exploration and Definition, (3) Demonstration and Validation, (4) Development, Production and Deployment, and (5) Operations and Support. Figure 2-2 illustrates how using the CDM supports three major aspects of the LCMIS milestone phases via data modeling.

Figure 2-2. CDM Support for LCMIS



2-4. How the CDM relates to the Defense Data Model (DDM)

The CDM provides a common data architecture for USACE. The DDM serves a parallel function for the Department of Defense. They both house metadata pertaining to unique business areas; therefore, this metadata may differ depending on specific USACE application requirements. Because the DDM is DoD's *Enterprise Data Model,* which contains an integrated view of data requirements common to all DoD functional areas, it is a USACE goal to align the CDM with the DDM.

A comparison is being conducted of the CDM against the DDM to determine: (1) the number of data elements in the CDM that already comply with DoD standards and the number that do not comply, (2) the number of Corps data elements that have been accepted as a DoD standard, and (3) the number of Corps data elements for which there is no DoD standard. Once the CDM data has been mapped/matched to DoD standards, the next step will be to integrate the data with the DDM.

2-5. Overview of the Data Modeling Methodology and Techniques

Modeling is a structured and graphical method of studying and documenting business activities and data elements required to support business requirements. The Integrated Computer-Aided Manufacturing Definition (IDEF) Modeling Techniques have been adopted by USACE as its standard modeling technique and methodology to be used during strategic planning, requirement analysis, and system development efforts. IDEF was issued as a Federal Information Processing Standards (FIPS) for integration definition. The IDEF0 methodology is used to define the requirements and specify the functions (or process) of a business. This model consists of diagrams, text, and a glossary, all of which are cross-referenced to each other. On the other hand, the IDEF1X methodology is used for data modeling. It represents the structure and semantics of information within an environment or system.

Data modeling is used throughout USACE to capture and define data requirements for business and information technology initiatives. Data modeling is also used as a means to define system requirements and solutions. In addition, data models are used to design a physical system and/or database that correlates the specific needs of a systems implementation project and the business objectives of an organization. Outside DoD, data modeling may be accomplished through the use of several methodologies (i.e. Information Engineering, Object Oriented Data Modeling, and Relational Data Modeling). IDEF1X is the standard DoD methodology for relational data modeling and database design. Appendix A provides more detailed information regarding IDEF Modeling.

USACE uses the IDEF0 and IDEF1X modeling techniques in the development of activity and data models to graphically and descriptively depict USACE's business processes and data requirements. The CDM was developed and is maintained using IDEF1X. The IDEF modeling technique provides:

- An effective means for displaying the current (AS-IS) and the planned (TO-BE) environment;
- A means to accomplish interaction and consensus between user groups as part of system development;
- Consistency, which facilitates data integration and data sharing, and;
- Audit trails and documentation during system conceptual design.

2-6. Overview of Standard Naming Conventions

Standard naming conventions are guidelines that specify how data entities and elements should be described and defined. Standard naming conventions result in an improvement in the validity, accuracy, and reliability of the data values stored in the CDM/CDD. The existence of proper standard naming conventions has important implications for the USACE community. Standard naming conventions facilitate data sharing, data consistency, and reduce duplication of data across the organization.

The purpose of naming conventions is to eliminate any miscommunication within an organization when referring to a specific data entity. Abbreviations and acronyms are misleading, defined differently, and used differently throughout the organization. In Table 2-1 below, the example illustrates how the acronym “DA” means different things to different components within or outside an organization. The impact of implementing the “DA” acronym without standardizing its full name, definition, and associated characteristics can be costly.

Table 2-1. Naming Convention Sample

Group I	Group II	Group III
Acronym: DA	**Acronym**: DA	**Acronym**: DA
Meaning: Data Administration	**Meaning**: Department of the Army	**Meaning**: District Attorney
Field Length: 19	**Field Length**: 16	**Field Length**: 17
Functional Area: Department of Defense	**Functional Area:** Army	**Functional Area:** State Government

Standard naming conventions result in the data integrity of the CDM/CDD as well as:

a) Fewer data value errors;

b) Greater consistency in the definition and use of data;

c) Less redundancy;

d) Improved access to data;

e) Clearer definition of data requirements, and;

f) Fewer complaints from functional users, particularly from those who have experienced persistent data error problems.

In addition, standard naming conventions include applying “class words” to the business area data model data elements. A class word is defined as a word in the name of a data element or attribute describing the category to which the data element belongs. Class words are used to facilitate a common understanding and consistent approach for identifying and documenting data entities and data elements. They establish the general structure and format of data in the domain for that data element (DODD 8320.1-M). Appendix C provides a standard listing of class words.

2.7 Three Schema Architecture

There are three types of schemas that are used to support various perspectives of the USACE data. The American National Standards Institute, Standards Planning and Requirements Committee (ANSI/SPARC) has developed, what is referred to as, the Three Schema (Level) Architecture.

- **Conceptual Schema**. The conceptual schema represents the logical view, or data administrator’s view, of the data requirement. This view is represented as a semantic (conceptual) model of the information that is stored about objects of interest to the functional area. This view is a single integrated definition of the data that is unbiased toward any single application of data and is independent of how the data is physically stored or accessed. A fully attributed, normalized data model is also referred to as a conceptual schema. The conceptual schema is used for data standardization and database design. It provides a consistent definition of the meanings and relationships of the data that is used to integrate, share, and manage the integrity of data within and across applications and user communities.

- **Physical (Internal) Schema.** The internal schema represents the physical view of database administrator’s view of the data requirement. This view is described by the data definition language (DDL) and physical storage methods used to implement the data requirements described under a conceptual schema.

- **External schema.** The external schema represents the user's view, or application view, of the data requirement. This view is represented by reports, transactions, and screens that are designed to support the individual worker, or groups of workers, in performance of tasks and activities. The external schema is often referred to as the end-user view.

2-8. The Linkage between Conceptual, Physical, and External Schemas

Conceptual models are data models that describe relationships among the entities and business rules of a business area. Conceptual models can be used to: (1) facilitate a common understanding of the data that supports organizational business functions or processes, and (2) serve as the basis for automated information systems (AIS) design and development.

When a systems development team plans to develop an AIS, the team must first develop a blueprint that serves as the framework for development of the system. The blueprint determines the size of the system, the data that will be contained in the system, and the interrelationships between the data contained in the database tables. Within USACE and DoD, the blueprint for the development of relational databases is the IDEF1X Data Model, also referred to as a conceptual model.

Conceptual models capture data structures in a format that is similar to the data definition language of a database system and are used to develop physical database structures for systems development. A physical database structure is referred to as a physical model. The physical model is a performance driven adaptation of the conceptual model where derived values may be used, entities may be combined or split into records, and data elements may be repeated in different records. The conceptual model relates to the physical model in the following ways:

- The conceptual model provides the standard metadata, including the fundamental data and definitions, which are the building blocks of the physical model.

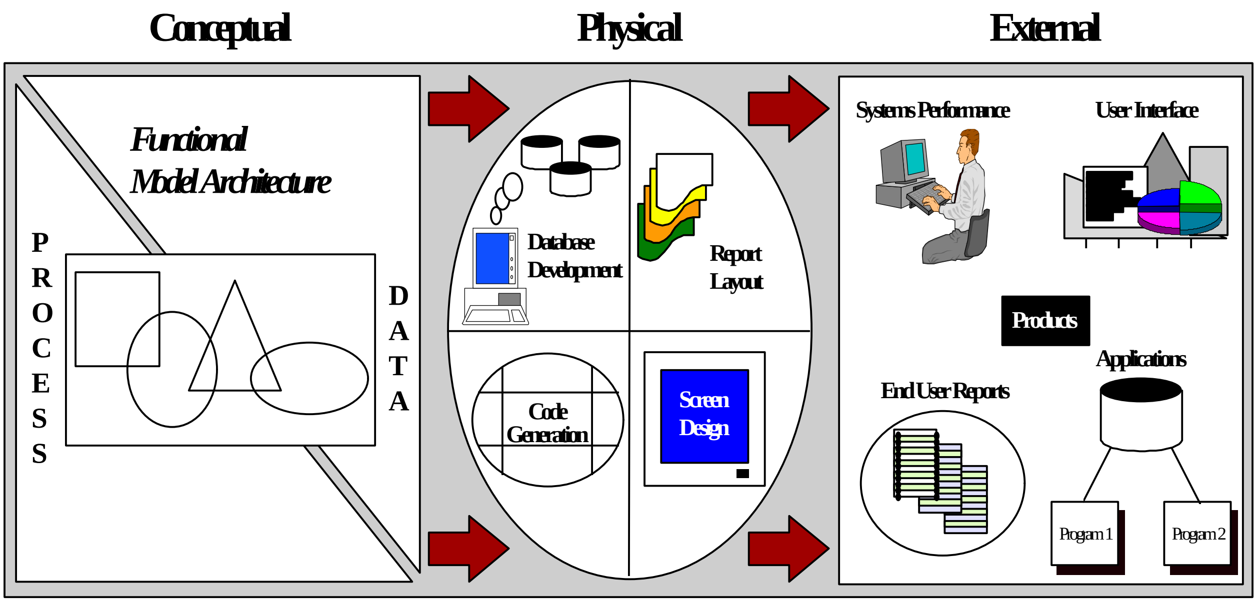
- Most of the entities in the conceptual model will become tables or records in the physical model.

- The fields of the physical model usually correspond to data elements in the conceptual model.

- Data element domains from the conceptual model are used to further define the physical fields in the physical model.

System developers must also map the data requirements of the system that are defined by the conceptual model to the external parameters of the end-user. These end-user requirements take the form of screens, reports, interactive sessions and forms that support the end-user in performing their functions. This "external" view is mapped to the physical model through a database management system (DBMS) and application programs. Figure 2-3 depicts how both the process and data models are used for database design and development and the linkage between all three schemas.

Figure 2-3. Conceptual to Physical to External Schema Correlation

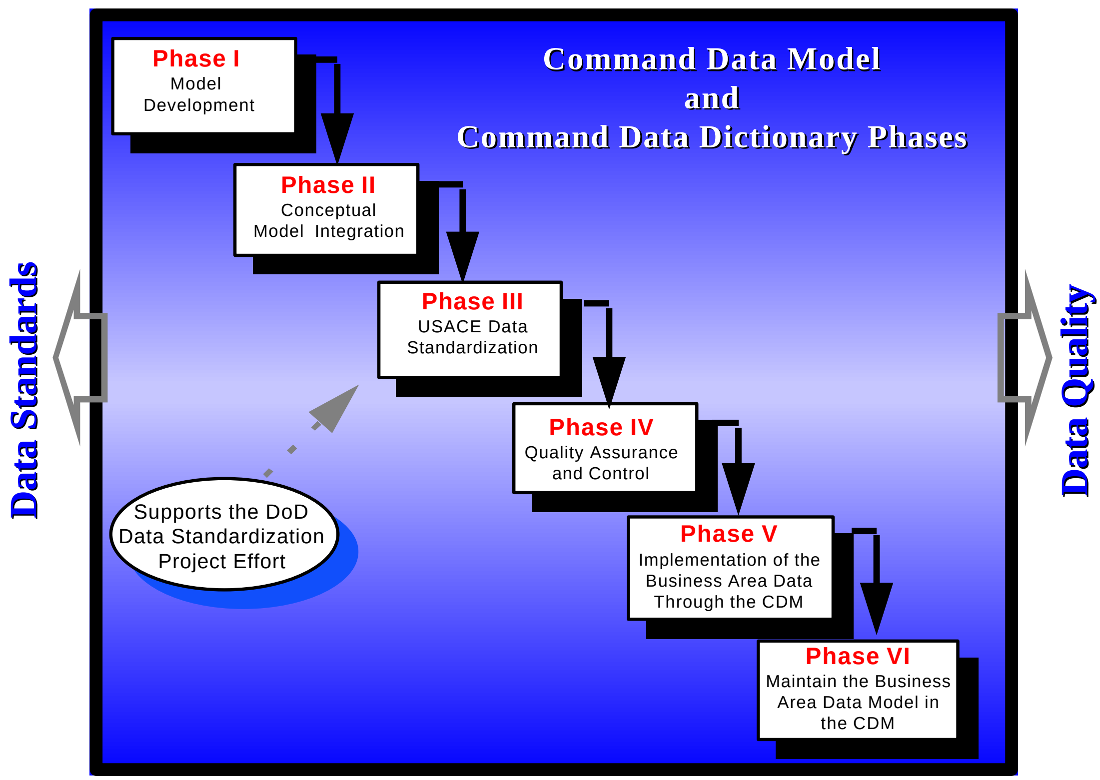


Chapter 3 - Standard Operating Procedures (SOP) for the CDM/CDD

3-1. Overview of Approach

This chapter outlines the Standard Operating Procedures (SOP) for using, maintaining, and implementing the CDM to support the USACE Data Administration Program. The procedures facilitate a common understanding of techniques and methodologies available to support data element standardization and data reuse USACE-wide. This chapter documents a phased approach that incorporates activities to ensure the validity, accuracy, and reliability of data for enterprise use. Figure 3-1 depicts the major phases in using, implementing, and maintaining the CDM/CDD.

Figure 3-1. Overview of the CDM/CDD Phases



3-2. CDM/CDD Phases

Provided below is a summary of the CDM/CDD phases. Detailed standard operating procedures for each phase are outlined in the following sections.

Phase I. Model Development - determining and documenting business and data requirements to support the organization's mission.

Phase II. Conceptual Model Integration - eliminating data redundancies and identifying data for reuse.

Phase III. USACE Data Standardization - developing standard naming conventions and definitions to facilitate data consistency.

Phase IV. Quality Assurance and Control - improving or restoring the quality of data in corporate information systems.

Phase V. Implementation of the Business Area Data Through the CDM - executing the USACE Enterprise Model as a framework for business and IT initiatives.

Phase VI. Maintain the Business Area Model in the CDM - ensuring that data is accurate and up-to-date in a timely manner.

3-3. Phase I. Model Development

a. Definition:

Model Development is defined as the process by which business activities and data requirements are graphically captured and documented. Facilitated group sessions are conducted with USACE management and subject matter experts to (1) fully comprehend the USACE business and (2) define pertinent data entities, subject databases, and interrelated business functions. Fully attributed data models are produced as a result of this assessment. This task includes enterprise data modeling using the USACE CDM. This will lead to the identification, process description, and prioritization of functional activities that need improvement. It results in a conceptual data model and a set of requirements for an information system to support the improved process and business rules.

The functional proponent initiates the first step in model development by identifying the needs and requirements of their organization.

b. Objectives:

The primary objectives of this phase are to:

- Analyze the existing business environment;

- Understand the business rules and how the business functions interrelate;

- Develop a balanced three-schema (conceptual, internal, and external) architecture framework;

- Develop a plan to guide improvement of the existing business processes and development of a supporting information system based on a three-schema architecture framework, and;

- Identify and prioritize conceptual modeling projects with a development plan.

c. Standard Operating Procedures:

Step 1. Establish the Data Modeling Scope

The scope of the data modeling business area is determined by reviewing the mission and objectives, and assessing the problem areas within the organization. The functional proponent is responsible for developing a charter that (1) states the problem and/or issues, (2) formally establishes the project objectives, (3) defines an initial team composition, and (4) outlines the functional proponents expectations of a project team.

The functional proponent should assemble a team to define initial project boundaries. The purpose for establishing boundaries is to:

- Clearly define the project scope to reduce the risk of gradual "scope creep";

- Establish the assumptions upon which the project will be planned and executed, and;

- Define a foundation for developing the remainder of the project plan.

At the conclusion of this step, the mission statement should be defined for the project. The mission may be refined in subsequent steps, at the discretion of the project functional proponent, as more information becomes available. The following information should also be identified and documented:

- The nature of the project;

- Management’s expectations;
- Initial constraints and risks;
- The environment in which the project will be executed, and;
- Preliminary success criteria.

Step 2. Identify and Assemble the Data Modeling Team

The functional proponent should identify a core team to develop a business area data model. Team members should be knowledgeable about the subject matter and generally include:

- The Project Manager (which may be the Functional Proponent);
- Subject Matter Experts;
- Key Technical Personnel, and/or;
- Data Administration Team Representative.

The data modeling project team is generally small, approximately 6-10 individuals. The scope of the project will dictate the team size. Additional resources may be called upon to provide input as needed. A data modeling project team usually consists of individuals with a variety of work backgrounds and skills which the functional proponent and/or Project Manager must build into a cohesive working group.

Capabilities and experience of potential team members are likely to include:

- Functional experience;
- Industry experience;
- Application design and/or development skills and experience;
- Analytical (Process/Data Modeling) skills;
- Specific technical expertise, and;
- Leadership skills and experience.

The functional proponent should be actively involved during the selection of the project team. If a team is formed and the mandatory level of technical ability, business knowledge, managerial focus and/or motivation is missing, the functional proponent should recognize this situation as a risk. Corrective action should be taken to reduce the risk to a manageable level. For example, if a functional proponent acquires a team that is deficient in technical ability, arranging training as soon as it is feasible would be an appropriate risk reduction measure. Support for data modeling projects can be acquired through the Directorate of Information Management, USACE Data Administration Office.

Step 3. Identify and Obtain Pertinent Documentation and Materials

The Project Manager should share background information and documentation with the team and should obtain their input on the preliminary document described in the preceding step. The data modeling project team should review existing documentation to gain an understanding of the project background and history. Relevant documentation may include, but is not limited to:

- Project proposal;
- Existing USACE process and data model constructs and model reports;
- Existing DoD data model constructs and data model reports;
- Project description;
- Organization charts, and;
- Existing functional and/or system documentation.

A description of the strategic objectives of the project should be described in one of the documents. If these objectives are not defined, the functional proponent or project manager should document the project's strategic objectives in preparation of the data modeling effort.

During the review process, the functional proponent should identify and evaluate:

- Implied commitments and concessions that have not been formalized or clearly documented;
- Constraints;
- Unclear and/or undocumented anticipated benefits of the project, and;
- Initial cost, resource or schedule estimates that appear to be unsubstantiated or potentially unrealistic.

Step 4. Develop a Workplan

The Project Manager is responsible for preparing a workplan for the business area data model project. The workplan should entail project task descriptions, estimated duration, roles and responsibilities, and a project schedule. The work plan should be presented to the functional proponent and Command Data Administrator for their review and approval. All parties must commit to an adequate number of man-hours to successfully complete the integration project.

Step 5. Conduct Data Modeling Training Workshop

Training should be given to the project team so that they understand the methodology and techniques that will be used throughout the duration of the project.

The objectives of the data modeling training workshop are to:

- Promote a common understanding of the IDEF methodology and modeling techniques

- Establish guidelines for the data model development, integration, standardization, and implementation activities

Step 6. Initiate the Development of a Conceptual Data Model

An IDEF0 Process Model should be developed as a basis for deriving data entities for the development of an information construct (i.e. conceptual data model.) This step directly supports the functional proponents with meeting Milestone 0 - Mission Need Justification and LCMIS Requirement. This process should outline the requirements to meet with mission and objectives of the target environment.

Forward and reverse engineering methodologies are used to develop conceptual data models. Forward engineering is the process of transforming an organization's business processes and data requirements (IDEF0- process model and the IDEF1X - data model) in a new physical design. This process involves interviewing subject matter experts, researching policy and procedure documents, and obtaining a full understanding of the organization's business area. The information resulting from the in-depth research and interviews will facilitate identification and understanding data requirements and business rules. Data requirements and business rules that are depicted in the form of a conceptual model are converted to a physical data model through the forward engineering process.

Converting an application function from one technology to another requires an understanding of the structure of information in the old system and the design of an information structure for the new system. Reverse Engineering is focused on the challenging task of understanding legacy program code without having suitable documentation. To be able to reverse engineer a data model, the design decisions essentially have to be reversed. Following the transformational approach, the forward engineering methodology is applied "backwards" to reverse engineer code to a more abstract specification (a conceptual model). A by-product of the transformational reverse engineering process is a design database for the program that can be maintained to minimize the need for further reverse engineering during the remaining lifetime of the system. The reverse engineering step may also be used simply as a way to obtain an understanding of the current system even if no revisions or conversions are currently planned.

Step 7. Develop Entity-Relationship (E-R) Information Construct

When information requirements of an organization are captured within an IDEF1X Data Model, the conceptual structure representing the organization's business area is captured in a normalized, non-redundant fashion. During the earliest stages of modeling, an Entity-Relationship (E-R) model is developed relating the major business entities together. It is during these early stages of development that a precedent is set for metadata reuse.

The practice of data reuse involves the sharing of standard business rules contained in the CDM/DDM enterprise models. The reuse of conceptual model data entities and elements is an iterative process, occurring throughout the modeling process. Reuse of the CDM and DDM metadata shall be applied where practical. From the onset of the Entity-Relationship (ER) diagram through the completion of the Fully-Attributed (FA) model, the reuse of CDM and DDM standards should be a consistent practice. This practice confronts integration issues at the forefront of the modeling effort by reusing existing or developing standards within the model's structure. This practice will serve to reduce the overall cost and time required for CDM integration and DoD standardization. The reuse of existing data standards from the CDM and DDM will also help promote interoperability among disparate information systems and reduce the presence of redundant data at the enterprise level. By reusing standards, the functional proponent ensures interoperability with standard objects, now and in the future.

Step 8. Develop Starter Model

Once the E-R information construct is developed, the functional proponent extracts data objects from the CDM and the DDM that are applicable to the functional proponent's business area to be used as a baseline "starter" model. The starter model is used for further evaluation to identify overlapping data and/or potential data reuse.

USACE currently has a tool suite that supports process and data modeling projects. This tool suite consists of the Data Encyclopedia, BPwin for process modeling, and ERwin for data modeling. The Data Encyclopedia and ERwin tools allow users to either build the physical data structure from the data model or reengineer a physical system into a conceptual data model. Recently an interface has been created between the ERwin and the Data Encyclopedia which allows users to either upload a data model that they created in ERwin into the Data Encyclopedia, or download a model from the Data Encyclopedia into ERwin for viewing purposes. Chapter 4 -

“Supporting Tools and Techniques,” provides an overview and benefits of each tool. These tools and techniques allow the Functional Proponent to develop their starter model either in the Data Encyclopedia and/or in ERwin. The steps in building a starter model are as follows: (See “USACE Data Encyclopedia End User Manual,” October 18, 1991 for more details)

(1) Query Data Objects in the CDM

The CDM is queried to identify existing data objects (entities, key and non-key data elements, domains, etc.) and applicable business rules which may also support similar information requirements within the proponent’s business area. Search and retrieval is supported through the Data Encyclopedia. The steps in querying CDM objects can be found in the “USACE Data Encyclopedia End User Manual,” October 18, 1991.

(2) Query Data Objects in the DDM

The DDM is queried to identify existing data objects which may also support similar information requirements within the proponent’s business area. The model developers must utilize the Defense Data Dictionary System (DDDS) or the Personal Computer Access Tool (PCAT) tool to locate suitable objects, and then trace their origin back to the Independent Entity Views (DDM) to analyze its structural context. The steps in querying data objects in the DDDS can be found in the “DDDS End User Manual,” and the steps in querying data objects in PCAT can be found in the “PCAT User’s Guide.”

(3) Extract CDM/DDM Data into a Starter Model

Based on the outcome of the queries of the CDM/DDM, any structures supporting the proponent’s business requirements should be extracted and copied into the business area data model. This extraction is referred to as a “starter” model which is to be used as a baseline for model development. The starter model should be refined by the project team, adding more detailed information about their business area. (See Step 9.).

Step 9. Evaluate Starter Model

Evaluating the starter model is a very critical task to the successful completion of the project. Any identified structures from the CDM and the DDM which support the proponent’s business area are evaluated against a predetermined set of criteria to determine its applicability in supporting the proponent’s functional area. This criteria should include, but not be limited to the following considerations:

- The level of normalization for the proposed structure and its ability to support the information requirements.

- The level of shareability of the business rule in either the Corps or DoD environment, now and in the future.

- The uniqueness of the data (if not located in the DDM, it should be standardized at the DoD level).

- Best fit of information structure semantics with the semantics of the proponent's business function.

- The ability of the standard objects to satisfy the functional proponent's business requirement and the Corps' mission.

Step 10. Validate Identified Standard Construct

The starter model is reviewed with the Subject Matter Experts (SMEs) to validate the constructs and their ability to support the functional business area information requirements. If the identified structures do not meet the criteria, conflicts should be identified and coordinated with the appropriate USACE Functional Proponent and/or the DoD FDad to resolve inconsistencies.

Step 11. Develop Detailed Information Constructs

This stage entails the iterative process of conceptual model development. The E-R model is transformed into a Key-Based (KB) model and then developed into a Fully-Attributed (FA) model. This process entails the Functional Proponent incorporating their unique data objects and business rules into the starter model, formalizing a fully-attributed business area data model. All required metadata, such as entity/element definitions, domain definitions/values and additional supporting information should be populated in accordance with FIPS PUB 184, Information Modeling Guidelines. Throughout model development, it is recommended to closely follow the DoD 8320.1-M-1 guidelines, especially those sections addressing entity and element naming conventions, definitions, and domain formats.

Step 12. Conduct Redundancy Analysis

The objective of this task is to evaluate the business area and data models to ensure that redundant data within the model is eliminated or justified. The project team should evaluate the model to ensure that it is in conformance with the Department of Defense (DoD) policy and guidance.

Upon the completion of the redundancy analysis, common names and rigorous definitions should be developed for all data entities and data elements. This in turn will lead to the elimination of duplication and incompatibilities in the data, processing, and dissemination of data.

Step 13. Conduct Model Normalization Analysis

The goal of data model normalization is to ensure that there is only one way to know a fact. Therefore, the technical process of normalizing a model removes all data objects that provide more than one way to know the same fact. Specifically, it is the process which places each attribute in an entity where it is dependent on the entire primary key of that entity and on no other key, as well as the elimination of data redundancies. This process is dependent upon a clear understanding of the data entities and attributes of the data model and the functional dependencies among the data objects in the data model.

The business area data model should be evaluated against the standard rules of data normalization. These rules will be used as the foundation for evaluating the degree of normalization of the business area data model. During this process, the project team should identify objects that represent business rules the organization that are not currently represented in the data model. Appendix F provides the normalization checklist.

Step 14. Load Model into the Data Encyclopedia

This step only applies if the Functional Proponent's conceptual data model has been developed in ERwin or other IDEF1X modeling tool. The business area model is loaded into a separate project within the Data Encyclopedia. The Data Encyclopedia is fully capable of supporting continued model development, and maintenance. The Data Encyclopedia also supports the ability to run data model correctness, completeness checks, and perform normalization to the third normal form.

Step 15. Conduct Trial Integration

Once the data model is completed and normalized, the project team should perform a trial integration to unveil potential risks and impacts against foreign-key migrations, parent-child relationships, cardinality, and category entities using the Data Encyclopedia. Entities, elements, and business relationships are evaluated for potential overlaps and assessed potential effects to a business area or information system. Detailed instructions on how to conduct trial integration on the Functional Proponent's business area model with the CDM, are referenced in the "USACE Data Encyclopedia End User Manual," October 18, 1991.

Step 16. Break The Model into Views

If the proponent's functional model is too large, it becomes necessary to break the model down into smaller subsets, usually grouped by logical views. This reduction in size will benefit future integration and standardization efforts. This process will assist the Functional Proponent with identifying the logical breakdown of their model as its business rules relate to the various business areas that the model may represent.

For formal DoD Standardization, it is mandatory to break down the model into manageable chunks, each representing one proposal package. The proposal package should generally contain no more than 20 entities and 200 attributes. When a data model is being developed that is larger than 20 entities and 200 attributes, it should be partitioned into separate views that can be submitted as individual proposal packages.

The USACE Data Encyclopedia will allow the partitioning of one model into separate views by creating subsets of the model into new models. This is easily accomplished by copying business rules into the new model. Once this has been completed, individual data dictionary reports can be generated. To maintain the integrity of the original model before it enters into any standardization efforts, it is recommended that a new project be created and that its newly devised logical subsets be created in this new project. This way, the original business information requirements are not altered by any modifications which may occur during the standardization process and which are not accepted in the final product.

3-4. Phase II. Conceptual Model Integration

a. Definition:

Conceptual model integration is defined as a consolidated abstract view of merged functional data models into the USACE enterprise-wide data model (CDM).

Conceptual Model Integration with the CDM is a team effort requiring the participation of the functional proponent, the USACE Data Administrator, subject matter experts, the data modeler(s) for the model being integrated, and the CDM integrator(s). Each integration will require different resources, depending on the size and complexity of the model being integrated, the degree of overlap with objects already in the CDM, the personnel assigned to the integration, and other factors. The following steps can be used as a guideline for a "typical" CDM integration. Each integration must be carefully planned and executed to ensure the integrity of the CDM so that it remains accurate, complete, up-to-date, and normalized (no data redundancies).

b. Objectives:

The primary objectives of this phase are to:

- Eliminate data redundancies;
- Promote reuse of standard USACE/DoD data entities and metadata;
- Enforce data consistency and integrity, and;
- Support systems integration and interoperability.

c. Requirements for Data Models, Prior to CDM Integration

Data models must meet the following criteria before they are integrated in the CDM:

- The model must be a fully attributed IDEF1X Conceptual Data Model, with all domains defined and attributed (including standard values).
- The model must reside in the USACE Data Encyclopedia.
- The model must pass the data model validation checks and the normalization checks.
- The model must conform to DoD 8320.1-M-1 regarding entity and element standards, such as class words, naming conventions, element formulation, domains, etc.

- The model must not contain any redundancies, acronyms, abbreviations or spaces.

- The model must pass the quality assurance checks as outlined in Chapter 3, Phase IV Quality Assurance and Control.

- The model must be up-to-date, and recently revalidated by subject matter experts.

- The model must be submitted to the Data Architecture Control Committee (DACC) for approval for "Registered" (shareable) status. In exception cases models may be integrated in "candidate" status prior to registration; the entities and elements in the CDM will be registered upon DACC approval.

d. Standard Operating Procedures:

Step 1. Prepare for Conceptual Model Integration

Conceptual model integration with the CDM is the responsibility of the USACE Data Administration Team, Directorate of Information Management. The functional proponent of a specific business area is responsible for briefing the Data Administration Team on the data model purpose, scope, contents, trial integration, and results of the data modeling project.

The Data Administration Team should obtain all documentation on the model and on the corresponding system (if one exists). Specific areas of interest include: (1) functions of the physical systems, (2) database structures, (3) database data dictionary, and (4) applicable regulations and circulars, etc. Evaluating these materials allows the Data Administration Team to become familiar with the business area processes and data requirements.

Step 2. Initiate Conceptual Model Integration

The business area data model is evaluated for quality, completeness, conformance to standards, and understandability. The business area data model quality assurance and control checklist for data model validation prior to integration. Appendix E provides the quality assurance and control checklist. If necessary, the model may have to be modified by the functional proponent in order to pass all checks prior to the start of integration. The data model completeness check must be executed using the Data Encyclopedia, and all errors need to be corrected by the functional proponent.

Based on the results of the model development phase, the Data Administration Team reviews the entity/element list, data dictionary, and data model relationship reports. All of the reports are compared against the CDM. The Data Administration Team will perform a trial integration of the models to unveil potential risks and impacts against foreign-key migrations, parent-child relationships, cardinality, and category entities using the Data Encyclopedia.

Step 4. Integrate with the CDM

During the integration task, overlaps between the business area data model and the CDM are identified and reconciled. This activity consists of (1) identifying the common entities, elements (including domain values), and relationships, and (2) determining their similarities and differences. The Data Administration Team is also responsible for identifying and documenting overlaps, resolving key and non-key migration issues, and providing support in developing common definitions for the shareable entities and elements. The functional proponents for each affected area will be consulted by the Data Administration Team for their approval of the proposed changes in writing.

The USACE conceptual model integration approach incorporates a thorough evaluation and mapping of all business area data model objects, such as entities, key and non-key elements, domains, relationships, and data definitions, against overlapping objects already in the CDM. This analysis differentiates data objects that can be reused as they are currently documented in the CDM, and those data objects which will likely have to be modified or merged. The more overlaps, the more complex or intricate the integration is likely to be.

The business area data model business rule report is annotated and used to show which relationships fall in the following categories:

- Already exist in the CDM;
- Conflict with those in the CDM;
- Need to be copied into the CDM, and;
- Require some change prior to being added to the CDM.

The Data Administration Team reviews the CDM and copies registered entities into the business area data model. After this analysis, changes are made to the business area data model entities and business rules. The team then analyzes the overlapping elements and domains ensuring that the registered data has already been reviewed thoroughly and requires a more intensive review prior to change. If a change is needed, the impacted proponent(s) should be consulted for agreement at this point. A formal registration review will be done later, but a formal functional approval (e.g. in person, by phone, email or fax) is required in advance to avoid problems.

The Data Administration Team will pay particular attention to standard domain values. When integrating two different lists, partial overlaps will be reviewed and resolved. At least two proponents will be consulted by the Data Administration Team to ensure that a combined list is compatible.

Step 5. Normalize the Shared CDM

The Data Administration Team will perform manual and automated normalization and correctness checks on the shared CDM after the business area data model is integrated. It is the final check on the newly integrated model to uncover data model errors. This will ensure that missing entity and element definitions, domains, key terms, duplicate data objects, and data modeling errors are evaluated and resolved. This task also involves the functional proponent's and business area subject matter expert(s') assistance to ensure that all data objects are defined and described accurately.

Step 6. QA Model

The Data Administration Team will conduct a thorough quality assessment of the model; annotating any question or deviation from standards or denormalizations. It is the responsibility of the functional proponent to make all changes to the model in order to pass the QA criteria. The data model must be complete and accurate prior to integration of data entities and elements into the CDM. Entities and business rules which can be reused as they are currently documented in the CDM must be reused. They should be integrated into the model with relationships. Only if these objects need to be modified should they be copied into the model as candidates. If modified, extensive notes should be taken on the changes made so that they can be evaluated and integrated later in Step 6 (Note: There is no "audit" to capture these changes automatically.)

Step 7. Produce Integrated Data Model and Reports

Once the business area data model is integrated with the CDM, a key term (usually the business area name, i.e. REMIS), will be attached to all applicable entities. This key term will be used to extract the integrated business area model from the CDM.

The Data Administration Team will produce a plot and/or other data model reports to ensure keys are still correct. The team will produce the final integration report and provide it for distribution to all DACC attendees. The report must be sent to the proponent first and then distributed for review by all integration team members. The integration report should only document the view being integrated (not all the views integrated to-date), except for the final integration report.

Step 7. Prepare CDM Engineering Change Proposal (ECP)

The Data Administration Team will prepare the CDM ECP for integration. Refer to Phase VI- "Maintain the Business Area Model in the CDM" for detailed instructions on completing the CDM ECP.

Step 8. Present to the Data Architecture Control Committee (DACC)

Prior to integration, the business area data model should be presented to the DACC for their review and approval. The DACC should be kept informed of all "requests for changes" needed to the CDM. All data entities should be presented to the DACC with emphasis on the data entities that are shared USACE-wide. The DACC should receive copies of the plan and of each integration report as each view is integrated with the CDM. (The steps for presenting models to the DACC can be found in the DACC Handbook.)

Step 9. Update Business Area Data Model Status

Upon DACC approval, the proposed data entities and data elements to be shared in the CDM will be registered. This change of status is executed through the use of the Data Encyclopedia. Detailed instructions on registering the data entities and elements is located in the "USACE Data Encyclopedia End User Manual," October 18, 1991.

3-5. Phase III. USACE Data Standardization

a. Definition:

Data standardization is defined as “the process of documenting, reviewing, and approving unique names, definitions, characteristics, and representations of data [entities and] elements according to established procedures and conventions.” (DoD 8320.1-M-1) The basic components of data standards include a conceptual data model and a standard data format.

b. Objectives:

Data standards have an enormous impact on how USACE shares data across functional areas. Program managers, functional proponents, and developers recognize the importance of data exchange and have established guiding principles to facilitate data integration, data standardization, and data sharing across the USACE. This capability demands technical support to ensure that data requirements are mapped and standardized across the enterprise in support of information management initiatives and the systems development life cycle process.

The objectives of USACE enforcing data standardization are:

- Support DoD Data Standardization Program in accordance with DoD Data Standardization Procedures, DoD 8320.1-M-1, November 1996.

- Promote standardization of data elements throughout USACE in a manner consistent with requirements for sharing data among USACE, DA and DoD.

- Enforce consistent data standardization guidelines throughout USACE during information quality assessments, data model development projects and systems design and development initiatives.

- Promote an interoperable environment where related systems are designed to share common data structures across the Corps and with external DoD systems.

- Reuse standard data elements to avoid duplication and non-standard data elements.

- Improve the manner in which USACE uses data by defining data structuring rules and standards.

- Improve coordination and modification of data entities and metadata.

- Train functional proponents in the use of the CDM and Data Encyclopedia and on developing standardization proposal packages.

c. Standard Operating Procedures:

The Standard Operating Procedures (SOP) for Data Standardization is provided in *detail* in the *Standard Operating Procedure for Corps of Engineers Data Standardization*, October 29, 1997. The phases below are a summary of the major activities to be performed.

Data standards are critical when developing, approving, implementing and maintaining systems within USACE information infrastructure. It is necessary to develop mission critical requirements in support of the functional proponents' business area. The collection of data should adhere to the USACE Data Standardization policy and guidance. When functionals develop and submit data standards to USACE or DoD the following procedures must be applied:

Step 1. Validate Data Entities and Metadata

As the project team validates the business area data model, the functional proponent must ensure that the model is in compliance with the USACE and DoD Data Standardization Procedures. The validation activities should include, but not be limited to, (1) ensuring that data entity/element names are in conformance with standard naming conventions, (2) verifying that definitions are structured accurately, and (3) ensuring that a complete listing of domain values for those data elements ending in the class word "code" is provided and defined.

Step 2. Determine the Relationship Between the Business Area Data Model and the DoD Data Model

In reviewing the USACE Data Model, the project team will determine the interface requirements to the DDM. When the project team discovers a direct match between the USACE Data Model and the DDM, the project team will specify the entity and its primary key from the DDM that has a relationship to a proposed entity in the USACE Data Model, and indicate where the USACE Data Model integrates into the DDM. Entities and their primary keys contained in the USACE Data Model "for display purposes only" must be in approved or candidate status in the DDDS.

Step 3. Coordinate Developmental Data Standards

The requirement to carry data through DoD Data Standardization is determined by the functional proponent who should ensure that coordination is done with other USACE business areas and the DoD Functional Data Administrator (FDAd). This review process should reveal the business impact of implementing a specific information structure and its benefits. The outcome of this review process is the resolution of conflicts with USACE functional proponents and DoD.

The functional model views are now submitted to the USACE Data Administration Team for their preliminary integration efforts with the Command Data Model. This step ensures that the DoD functionally related metadata is consistent and compliant with the business requirements of the USACE as depicted through the Command Data Model. If areas of contention arise, final resolution requires the collaborative efforts of the originating USACE functional proponent, the affected DoD FDAd, and proper Corps representation. This review process will ensure that all of the proposed standards originating from the Corps has both functional and corporate support prior to formal DoD submission.

Data standards originating in support of an Office of the Secretary of Defense (OSD) functional area requirement should be coordinated with the appropriate FDAd. Data standards originating within a Component or at the Component level should be coordinated with the appropriate Component Data Administrators (CDAds) and FDAds prior to entry into the DDDS, per reference DoD 8320.1-M-1.

It is the responsibility of the functional proponent to ensure that all data modeling products in support of USACE are submitted to the appropriate Functional Data Administrator (functional data steward(s)) and/or CDAd for review. The functional proponent may assist in coordinating a preliminary review of the USACE developmental data standards within the DoD functional community and resolve any issues identified during this preliminary review. This is an iterative process requiring the participation of the USACE functional proponent, USACE Command Data Administrator, Subject Mater Expert SME(s), DoD CDAd(s),and DoD FDAd(s). Discrepancies and major issues should be documented and reviewed by the business area functional proponent at the USACE and DoD functional levels.

The FDAd at the DoD levels are consulted when the functional proponent is in the process of DoD Data Standardization. This review process should reveal the impact of implementing a specific information structure and its benefits. The outcome of this review process is conflict resolution with USACE functional proponents and DoD.

Step 4. Develop USACE Data Standards Proposal Packages

The project team will incorporate the business area data model into a DoD data standards proposal package ready for submission for formal review and approval under DoD 8320.1-M-1 series guidance. “The FDAd will oversee the assembly of a package which proposes the functionally coordinated developmental data standards as an extension or update to the DDM.”

For formal DoD Data Standardization, the business area data model must be presented in a consistent and practical manner. This is accomplished with the creation of a proposal package which can be independently processed. It is usually necessary to break down the model into manageable sections, each representing one proposal package. The proposal package should generally contain no more than 20 entities and 200 attributes. When a data model is being developed that is larger than 20 entities and 200 attributes, it should be partitioned into separate views that can be submitted as individual proposal packages.

All data models must be submitted in IDEF1X format in compliance with FIPS PUB 184 and DoD 8320.1 series guidance. The functional proponent is responsible for making any corrections, additions, deletions or modifications identified as a result of the technical and functional review process in the Data Encyclopedia.

Appendix D provides a data standardization checklist in preparation for proposal package submittal to DoD.

Step 5. Conduct Final QA of the Metadata

A final QA of the USACE metadata is conducted by the functional proponent prior to submitting the proposal packages to Defense Information Systems Agency (DISA) to ensure that all the metadata is complete, and in compliance with the USACE and DoD Data Standardization Procedures. It is highly recommended that all data models are normalized at least to third normal form. As an additional reference, see “Standard Operating Procedures for U.S. Corps of Engineers Data Standardization” (draft).

3-6. Phase IV. Quality Assurance and Control

a. Definition

Quality Control is defined as an aggregate of activities (such as design analysis and statistical sampling with inspection for defects) designed to ensure quality. Therefore, the purpose of quality assurance and control is to ensure that the deliverables of a project, such as data models and project documentation, are error free, up-to-date, and satisfy the requirements of the project.

b. Objectives

The role of quality assurance has several objectives when applied to using, implementing, and maintaining the CDM:

- Ensure the project methodologies, standards, and guidelines identified in this publication are implemented throughout the duration of the project

- Ensure the specific standards and measures identified in detailed work plans are appropriate and reasonable for achieving the defined quality objectives for specific deliverables and data models

- Ensure an appropriate quality training program for all project participants is developed and implemented (i.e., IV&V of Data Models)

- Review for adequacy and adherence to general project controls, such as technical methodologies and change control procedures

- Ensure quality control activities are performed consistently and continuously

- Analyze errors and defects to provide a basis for improving the quality of project data models and data model reports/documentation

c. Standard Operating Procedures

Once the business area data model has been developed, there are a sequence of steps that need to be performed to ensure its integrity. The steps are as follows:

Step 1. Plan for Quality

The achievement of quality must be planned. Therefore, the first step in managing quality for data modeling projects is to develop a plan to achieve project goals. The objective of this step is to ensure all the standards and guidelines required to effectively plan, manage, staff, control, and deliver the project have been identified.

Planning for quality begins when project level standards are established during the management planning process. Review and sign off requirements should be defined explicitly for individual deliverables. The functional proponent should ensure standards and guidelines adopted for the specific project are both sufficient and appropriate.

Step 2. Establish Quality Assurance Framework

To be successful, the quality assurance framework must have strong management support. This support must come from the functional proponent, project manager, and all other levels of management involved with the project effort. Management support must be threefold:

- The quality function must be adequately staffed with skilled resources;

- The quality function must be given the authority necessary to fulfill their responsibilities, and;

- Management must be supportive of recommendations.

In addition, the functional proponent must assume responsibility for reviewing all significant findings and recommendations and acting upon the suggested actions in a timely manner. The functional proponent must ensure that the detail work plans are continually updated to reflect implemented recommendations.

Step 3. Perform Quality Control Activities

Quality control should be performed to identify and correct defects in project deliverables and documentation. Quality control is the responsibility of each project team member and ultimately the functional proponent. Quality control must occur throughout the project, and not just when a deliverable is complete. The three methods generally used to monitor and control quality are: Reviews, Independent Verification/Validation, and Surveys/Questionnaires.

(1) Reviews

The general purposes of reviews are:

- Verify that all data models and deliverables conform to specified standards and guidelines.

- Ensure consistency among deliverables.

- Provide a forum to evaluate the content and correctness of the deliverables.

- Provide a method to verify and evaluate objectivity, clarity, correctness, and completeness of all deliverables, documents, and products.

However, reviews can easily become more subjective than objective if not managed effectively. The Project Manager or functional proponent must clearly define the objectives of the review in accordance with the following:

Completeness -	No items are overlooked or partially completed;
Consistency -	The work is consistent with other project components;
Conformance -	The work adheres to USACE preestablished data model project methodologies, standards, and guidelines such as IDEF;
Objective -	The work meets the task objectives and adheres to specifications and satisfies the functional proponent expectations and requirements;
Clarity -	The work is unambiguous and comprehensible.

There are a number of different review methods that may be effective for projects. A successful quality approach generally includes a combination of review methods. These review methods encompass periodic checks conducted by the members of the project team, as well as by the functional proponent. The project team assigned to produce the deliverable must compare attributes of the deliverable against predefined criteria or a checklist. Appendix E and Appendix F provide comprehensive checklists to validate data models and to ensure that proper coordination has been done prior to project completion. A checklist may identify missing or incomplete components of the deliverable.

(2) Independent Verification and Validation (IV&V)

Verification and validation by an independent source is the second primary quality control technique used to determine and control the quality of deliverables. This activity is used to compare and verify actual results against expected results or established standards in an objective manner. Therefore, IV&V objectives should be defined clearly and as early as possible in the project life cycle. Comparisons between expected and actual test results are generally quantified and classified to highlight strengths and weaknesses.

IV&V does not only apply to models, but to all aspects of the project. IV&V activities should be used as a quality control for:

- Activity and Data Models;
- Functional Requirement Specifications;
- Work Plan;
- Training Materials, and;
- Reports/Documentation.

This approach focuses on performing functional and technical verification of products. Functional verification is the mapping of functional specifications against the functional proponent's needs and business requirements. Technical verification ensures that the data models meet all technical guidelines and requirements in accordance with the FIPS 184 and USACE Data Administration policies and procedures. Appendix G provides a checklist of IV&V activities.

(3) Surveys and Questionnaires

A third quality control method is surveys and questionnaires. Surveys are particularly valuable when conducted during a project to measure functional proponent/customer satisfaction with the project approach, processes, and techniques, as well as a measurement of satisfaction with the end product. Project managers are expected to measure, on a regular basis, client satisfaction with services provided.

Step 4. Implement Corrective Action

Defects and/or non-conformities discovered in a data model or deliverable arising from a quality control checkpoint must be corrected. To effectively correct the problem, the root cause of the defect or nonconformity should be determined. For effective corrective action, maintaining good quality records is essential. Comprehensive records enable the Project Manager to track and identify the source of defects. Records are necessary to demonstrate quality control points that have been established and working as planned. These records should be current, comprehensive and readily available.

Quality efforts necessary for implementing corrective actions include the following:

- Identify and categorize the defect and nonconformities to determine if problems are unique (one time) or recurring throughout the model or deliverable;

- Identify realistic corrective actions and select most cost effective solutions;

- Implement the solution and measure the results; and,

- Determine if achieved results are commensurate with anticipated results.

Managing quality is a continuous process. As data is analyzed, it may become obvious that changes in standards or guidelines are required. Quality control methods may also require revision if errors are being missed in early project phases. The quality assurance framework may need to be changed or updated as quality control data is analyzed and corrective actions identified. The functional proponent should anticipate constant refining of the quality process.

3-7. Phase V. Implementation of the Business Area Data Through the CDM

a. Definition

Implementation of the business area data is defined as the process of designing and implementing the data structures for the AIS by executing the business rules, standard definitions, and supporting metadata integrated within the CDM.

b. Objectives

This section explains the critical role of the CDM in the development cycle of Automated Information Systems and defines procedures associated with AIS development. Automated Information Systems development begins with the conceptual model, proceeds to the physical model, and then to the external model. The conceptual model is composed of entities, attributes, and relationships, and is independent of any applications derived from this data. The physical model is composed of the programming and relational databases, including tables/fields associated with the development of business specific automated information systems, while the external model presents the final users' view (screens). This is the ***System Development Life Cycle*** whose architectural foundation is the three-schema data architecture (see paragraph 2-7). However, the CDM represents only the aggregate of the conceptual models and not the physical and external models. The objective of this section then, is to define and disseminate the procedures associated with the transition from a DACC approved conceptual model to the ***implementation*** of the physical and external models in an AIS.

c. Standard Operating Procedures

The stated goal of the Information Technology Strategic Plan for IM Master Strategy of the USACE Strategic Vision calls for revolutionizing effectiveness by "bold process reengineering and innovative use of technology". Putting in place the following standard methodologies for systems development will greatly assist USACE in realizing this goal.

In accordance with the Life Cycle Management Information System (LCMIS), it is a stated USACE policy that all new AIS development, using shared data, will be incorporated into the CDM as it relates to the business area. Therefore, it is the responsibility of the functional proponents of new systems to ensure that their efforts are coordinated through the Data Administration Office. It is the responsibility of the Data Administration Team to provide oversight program management and technical support to the functional proponents in the use of the CDM over the course of the System Development Life Cycle.

To comply with LCMIS, the functional proponent should develop the conceptual, physical, and external views, and provide the conceptual and physical models to the Data Administration Team for inclusion into the CDM. After completion of the implementation of a new AIS it is the responsibility of the functional proponents to coordinate with the Data Administration Office, all system changes which would affect the CDM or Data Dictionary.

Following are the steps that the functional proponent should take in the AIS development cycle:

Step 1. Transform the Business Area Data Model into a Physical Model

1. Use the fully attributed data model to develop logical transactions statements. These statements represent discrete units of work performed against the database and are similar to the description of the lowest level processes in the model.

2. Develop the transformation model. This is a model of the physical database design and includes system control and access information as well as the business data from the fully attributed model.

3. Develop the database create statements using data definition language. Each entity in the transformation model becomes a table, and each attribute become a column or field.

4. Design the screens and reports for the external model.

Step 2. Upload the Physical Model into the Data Encyclopedia

The physical model is uploaded into the Data Encyclopedia after each new release of the physical system; more frequently if updates are made between releases. Some physical models are uploaded monthly. When a new physical model is uploaded, the physical model version difference report should be executed, and the report should be examined for changes to the physical system that affect/impact the data model, correlations, etc. This assessment also helps to ensure that the business needs, data model, and physical system are synchronized.

For those models with 3-schema views, follow a parallel procedure for external views.

Step 3. Maintain Conceptual to Physical Correlations

The project team should correlate the conceptual data model objects to the physical model objects. This activity entails the team correlating the data entities to database tables and data elements to physical field names. This activity assists the application developer and modeler with understanding how the conceptual business area model is implemented in a physical system. As system changes arise, correlations play an important role. Prior to implementation into a system, impact analysis should be conducted on those conceptual data objects that correlate to the proposed changes in order to assess the impact those changes may have on other business areas. It is vital to run, after each physical model upload, the "migcorse" program which migrates the correlations to the new version. The project team should contact the Program Manager of the Data Encyclopedia for access to the migcorse program.

Detailed instructions on how to correlate conceptual to physical models is located in the “USACE Data Encyclopedia End User Manual,” October 18, 1991.

Step 4. Maintain External Views

Metadata about reports, application screens, and paper forms utilized and produced by AIS’ in USACE are referred to as external views. External views are composed of fields which are known as external fields. External fields are used to display and input data for an application.

The project team should populate external view information into the Data Encyclopedia. Using the Data Encyclopedia, the external view information is stored and maintained. Information such as the application name, screen number, report name, or screen title should be stored in the Data Encyclopedia.

In addition, maintaining the external views correlations play an important role as system changes occur. Prior to implementation of a system, an impact analysis should be conducted on those screens, reports, and documentation that correlate to the proposed changes in order to assess the impact those changes may have on other business areas.

Detailed instructions on the external views and the external view correlations are located in the “USACE Data Encyclopedia End User Manual,” October 18, 1991.

3-8. Phase VI. Maintain the Business Area Model in the CDM

a. Definition

Maintaining the business area data model refers to those activities, performed after the initial integration of the business area data model into the CDM, which are required to ensure that the CDM reflects the most current version data for a specific business area.

b. Objectives

The purpose of this section is to impart the critical nature of the maintenance functions and describe the procedures associated with this function. Implicit in the nature of systems is change because systems reflect business operations which are not static. To maximize the value of the CDM to the USACE, the model must reflect these changes on an ongoing basis. The result of not maintaining the conceptual framework after the information systems are developed and later modified is a loss of business rules. This may result in a costly and time consuming remodeling activity. Failure to maintain the business area data models integrated with the CDM will degrade the value of the CDM as a data administration tool. Eventually that degradation will result in obsolete data, forcing a major, avoidable remodeling effort.

c. Standard Operating Procedures

The following is a summary of the steps used for maintaining a business area model. These steps are used for "incremental" changes, i.e. adding an element, changing an element length or type, adding or changing business rules, submitting changes to the DACC, and performing impact analysis of the system.

Step 1. Appoint Designated Organization

All business areas must have designated functional proponents, Automated Information Systems Data Administrators (AIS DAd) and Subject Matter Experts (SME) who are responsible for maintaining their conceptual, physical, and external views. It is the responsibility of the functional proponent to perform a change impact analysis (an economical and technical assessment of the impact of changes needed to comply with a requested modification) on cross functional system data. However, it is the responsibility of the Data Administration Team to analyze the impact on the CDM.

Step 2. Identify Data Changes

New or updated data/business requirements are reviewed and documented in a form of an Engineering Change Proposal. Any change to an AIS or business area data structure must be included in the business area data model before the changes can be implemented in a physical AIS or business area. All changes must be approved by the functional proponent and reviewed by the DACC prior to including the change in any released version of the AIS. All candidate changes must be elevated from registered status to approved status in the data model prior to field release of the AIS.

Step 3. Complete ECP

To submit a new data model change request, an ECP Form should be completed, indicating the data model changes, impact, and general information such as the originator's Name, Phone, E-mail, Model Name, Version, etc. Appendix H provides an ECP Form.

Step 4. Review the ECP

The originator of the ECP reviews the ECP with assistance from the AIS DA, AIS DBA, and the functional proponent. The AIS data modeler logs all changes which can impact the data model on an ECP form, giving it a unique ECP number.

Step 5. Submit Impact Analysis Changes to USACE Data Administration Team

The AIS modeler performs impact analysis on the change and notifies any functional proponents that is impacted by the change, and then submits the change to the USACE Data Administration Team for quality assurance.

Step 6. Integrate Change(s) into the Business Area Model

The USACE Data Administration Team reviews all technical aspects of the candidate change, including naming conventions, definitions, domains, relationships, etc., performs integration analysis by identifying any overlaps throughout the CDM, and provides feedback to the AIS modeler. The USACE Data Administration Team is responsible for producing an integrated solution.

When the AIS modeler and the Data Administration Team reach an agreement on the change, the change can be implemented in the physical database. The AIS data modeler makes the physical/conceptual correlation in the Data Encyclopedia.

Step 7. Prepare a Summary Matrix Log Report

The USACE Data Administration Team integrates into the CDM the change as a candidate, and a CDM ECP is completed, given a number, and logged. The Data Administration Team prepares a summary matrix log of all candidate changes made to the CDM since the last DACC meeting, and provides an electronic copy to all functional proponents.

Step 8. Update Candidate Status

Once the DACC approves the proposed candidate changes outlined in the summary matrix, the entities and data elements will be registered. This change is executed through the use of the Data Encyclopedia. Detailed instructions on registering the data objects is located in the "USACE Data Encyclopedia End User Manual," October 18, 1991.

3.9 CDM Quality Assurance (QA) Maintenance Procedures

The following is a summary of the steps used by the Data Administration Team in preforming quality assurance of the CDM. These steps are used for preparing change requests, changing status codes of data objects, running various reports, and updating the counts of the CDM objects.

Step 1. Prepare CDM ECP Form

A log of all CDM changes must be kept and recorded for each change documented on an Engineering Change Proposal (ECP). The CDM QA person reviews the ECP and files it in the CDM Management Control Log binder. The functional proponents signature is checked.

Step 2. Post CDM Summary Log

Two weeks prior to each DACC meeting, CDM personnel assist the DACC Secretary in producing the CDM Summary log for distribution to the DACC. The Summary Log is also posted to the Internet.

Step 3. Run "Candidates and Registered Entities" Report

Every two months the report, "Candidates and Registered Entities," is run, and the items that the DACC have been notified of through the CDM Summary Log are updated from candidate to registered.

Step 4. Distribute Copies of CDM Disk

Prior to every DACC, the DACC Secretary produces copies of the CDM disk, and distributes them to each functional proponent.

Step 5. Run Normalization Report

After integration of each new data model, a normalization report is run via the Data Encyclopedia, the diagnostics section is analyzed, and corrections are made to the CDM.

Step 6. Run "Data Model Correctness Check" Report

After integration of each new data model, the data model correctness check is run, and all errors/warnings are corrected.

Step 7. Run Report of CDM Data Dictionary

Every six months an archive/reference of the CDM data dictionary is run, bound, and labeled.

Step 8. Scan CDM for Update of Object Count

At least once a month, a scan of the CDM is made, and a count comparison of entities, elements, and business rules are tabulated to the prior CDM to ensure that no data is lost.

Appendix G provides the model development process checklist to ensure that all the pertinent steps described in this SOP have been implemented.

Chapter 4 - Supporting Tools and Techniques

Common procedures and automated tools are used and continually improved to support functional managers, data administrators, and the technical support community in providing, maintaining, and using standard data products.

USACE has established guidelines and automated tools to support all components of the data administration program. Common tools are essential to use the CDM/CDM and facilitate the data administration activities by facilitating the reuse of existing data objects and resources and providing guidance to functional proponents on modeling their data architecture requirements. This includes conceptual models and standard naming conventions for forms, records, and automated information systems as previously mentioned.

To date, there are a number of Computer-Aided Software Engineering (CASE) tools that support full systems development life-cycle, from analysis to code generation. The later system development phases, such as system implementation and physical database design are supported by only a few, highly specialized tools. Linking methodologies and tools currently being used throughout USACE is essential to the successful development of the data administration infrastructure. This will aid in the development and modernization of USACE's automated and non-automated information systems. Components should attempt to migrate to USACE-wide (and ultimately, DoD-wide) common methodologies and sets of tools (e.g., information engineering and CASE tools) that conform to USACE's criteria as closely as possible.

At a minimum, the USACE uses such tools as ERwin/BPwin, the Data Encyclopedia, and the Defense Data Dictionary (DDDS).

4-1. Overview of the Data Encyclopedia

The purpose of the USACE Data Encyclopedia (referred to as "Encyclopedia") is to support and maintain the CDM. The Encyclopedia was developed under the Information Engineering, Data Administration, Database Administration, and Application Development Project sponsored by the Information System Modernization Program (ISMP).

The Encyclopedia provides USACE with an automated integrated repository for metadata which facilitates shared access to data/information by users. It contains metadata about conceptual data objects, physical data objects, and external view and application objects.

The Encyclopedia provides usable benefits in the area of activity and data modeling, multi-graphical interfaces, and user-shared repositories of metadata. The Encyclopedia can assist the application developer in defining the data requirements of new or redesigned application software through the use of activity and data models.

The data repository can directly be queried regarding data models already in existence, thereby permitting the user to select a starter or seed model to identify data objects for reuse.

Detailed information regarding the encyclopedia functionality is documented in the "USACE Data Encyclopedia End User Manual," October 18, 1991.

4-2. Overview of ERwin and BPwin

Both automated and publication tools are needed to speed development and to manage and control the development process of building systems. Many CASE tools are available to help with the design standards that are needed to ensure consistent system solutions.

ERwin and BPwin are interactive modeling tools that operate under Microsoft Windows in an 80x86-based computer. They are more than a drawing package; the tools enforce the rules of IDEF0 and IDEFIX in the construction of diagrams. They provide a full range of reports and diagram views. ERwin is used in conjunction with IDEFIX to model data, and BPwin is used with IDEF0 to model functions or processes also referred to as Activity Modeling.

4-3. Overview of the Defense Data Dictionary System (DDDS)

The DDDS is the primary automated tool that supports DoD Data Administration in developing and managing standard data per Directive 8320.1. It is the DoD Data Administration's Information Resource Dictionary System which provides a centralized repository of information about standard data, accessible to DoD users.

The tool has automatic message capabilities and is designed to permit the approval of DoD standard elements in a timely manner. Accessibility of the DDDS to users throughout DoD permits reduced data redundancy and increased sharing of accurate information.

4-4. Overview of the Personal Computer Access Tool (PCAT)

PCAT is a standalone version of the DDDS. PCAT provides a user friendly method of searching and analyzing all the DoD standard and nonstandard data elements and prime words that reside in the DDDS. It provides a mechanism for defining metadata, cross-referencing and consistency checking, and supports the standardization of data element names, definitions, and relationships. PCAT was developed to support the requirements of data analysis and data administration audiences.